

Original Article

Transcriptome profile and pathway analysis of starch and sucrose metabolism in *Euglena gracilis*Wenhui Zhang^{1,2}, Wen Jye Mok¹, Jinwei Gao², Yeong Yik Sung¹, Wenli Zhou^{2*}¹Universiti Malaysia Terengganu, Institute of Marine Biotechnology (IMB),
Kuala Terengganu, Terengganu, 21030 Malaysia²Tianjin Key Laboratory of Aqua-Ecology and Aquaculture, Fisheries College,
Tianjin Agricultural University, Tianjin, 300384 China

Received: 9 April 2021; Revised: 20 March 2022; Accepted: 5 May 2022

Abstract

Euglena gracilis, a single cell flagellate eukaryote, produces secondary metabolites known as paramylon, a carbohydrate similar to starch with beneficial health properties. In this study, we sequenced the transcriptome and identified the pathways and key enzymes involved in the starch and sucrose metabolism of *E. gracilis*. A total of 120,086 unigenes were assembled with the Trinity software, of which 48,031 were annotated. The genes involved in the starch and sucrose metabolism of *E. gracilis* were further identified, leading to the construction of a preliminary pathway map of starch metabolism for this single cell flagellate eukaryote species. Both β -*amy* and *Spase2* were identified to play major roles in starch catabolism, whereas glucose-1-phosphate adenylyltransferase functions in starch anabolism. It is anticipated that the starch content of *E. gracilis* can either be augmented by inhibiting the starch breakdown with β -*amy* or *Spase2* gene interference or by elevating the level of AGPase activity of Glucose-1-phosphate transferase to promote starch anabolism. Whatever the outcomes, results from this study provided a fundamental understanding on how *E. gracilis* regulates its starch metabolic pathways, and the impact of genetic modification on starch synthesis.

Keywords: *Euglena gracilis*, high-throughput sequencing, transcriptome, starch and sucrose metabolism

1. Introduction

Euglena gracilis is a spindle-shaped, small (100 μ m long) freshwater unicellular eukaryote that belongs to the phylum Euglenophyta, class Euglenophyceae. The species was discovered and termed by the Dutch biologist Antonie van Leeuwenhoek in the 17th century. *E. gracilis* generates food through photosynthesis by carbon fixation (Ahmadinejad, Dagan, & Martin, 2007; Pulz & Gross, 2004), and acquires organic substances through osmotrophy (Afiukwa & Ogbonna, 2017; Ogbonna, 2009; Pringsheim & Hovasse, 2010; Takeyama *et al.*, 1997). This so-called “dual feature of plant and animal” enables *E. gracilis* to be highly rich in nutrients, including polysaccharides, vitamins, amino

acids, and unsaturated fatty acids (Zhang *et al.*, 2020). *E. gracilis* has been shown to accumulate large amount of polysaccharides under aerobic conditions, with β -1,3-glucan referred to as paramylon, being the major ones (Yoshida *et al.*, 2016). Paramylon are structurally similar to starch and they display potent anti-oxidative and immune-regulatory activities (Russo *et al.*, 2016). Sucrose and starch metabolism in plants involve several complicated biochemical and physiological processes, which is important because it determines the partitioning of assimilated carbon, but also because it may limit the entire photosynthetic process under some environmental conditions (Sharkey, *et al.*, 1986). Compared with other algae, *Euglena* does not produce a typical α -1,4/6-glucan, such as starch. Instead, when grown aerobically in light it produces an insoluble β -1,3-glucan. This glycopolymer shows immune stimulatory properties and is reported to have anti-HIV8 effects (O'Neill *et al.*, 2015). In this context, it is therefore important to understand the molecular mechanisms

*Corresponding author

Email address: wenlizhou@tjau.edu.cn

and synthetic pathways involved in starch and sucrose acquisition and production in this aquatic organism for future commercial and medical applications.

To date, many algal transcriptomes were retrieved to determine their transcript classification and relevant pathways, with *Dunaliella viridis* (Zhu *et al.*, 2015) and *Scenedesmus sp.* as examples, the latter aimed at revealing genes related to the carotenoid biosynthesis (Huang *et al.*, 2018). Additionally, the transcriptome of *Rhodomonas sp.* were used to determine the anabolic pathways of polyunsaturated fatty acids (EPA/DHA) (Li *et al.*, 2016) while those from *Chlorella minutissima* revealed genes involved in the biosynthesis and metabolism of triacylglycerol, the most abundant energy-dense storage compounds occur in eukaryotes, with their metabolism primarily important in lipid homeostasis, cellular energy balance, maintenance and growth (Yu *et al.*, 2016).

In this study, the transcriptome profiles of *E. gracilis* were used to unravel the starch and sucrose metabolic pathways. High-throughput sequencing with the Illumina HiSeq 4000 sequencing platform, a method widely used in determining the transcriptome profiles of plants (Muriira, Xu, Muchugi, Xu, & Liu, 2015; Yan, Qin, Feng, Zhou, & Fang, 2017) as well as microorganisms (Cloonan *et al.*, 2008; Nijkamp *et al.*, 2012; Ugrappa *et al.*, 2008), was applied to construct the transcriptome library of *E. gracilis*. The sequenced and assembled unigenes were classified, functionally annotated, and subjected to pathway analyses. Specifically, paramylon was indicated as a specific component of *E. gracilis*, and the transcriptome data revealed several key genes involved in the starch and sucrose metabolic pathways. Information presented herein provided a scientific basis for further research towards developing potential strategies for the direct synthesis of starch and sucrose metabolism from *E. gracilis*.

2.1 Materials and Methods

2.1 *E. gracilis* culture

E. gracilis FACHB-849 provided by the Freshwater Algae Culture Collection of the Institute of Hydrobiology, Chinese Academy of Sciences were inoculated in HUT and AF-6 culture medium, the latter prepared and sterilized 20 min at 121.3°C with a pressure of 1.05 kg/cm² before use. Temperature was maintained at 25°C ± 1°C. The initial density was 2.83×10⁵ cells/mL. In order to obtain more information, *E. gracilis* under four culture conditions were selected for sequencing. The culture flasks were vortexed several times a day to prevent permanent adherence and settlement of the microalgae.

2.2 RNA extraction and library construction

One hundred microliters of the 5 days culture solution were collected and centrifuged at 4°C (2,000 rpm, 10 min). The supernatants were discarded, and the samples were stored at -80°C. Total RNA was extracted using a Total RNA Extractor (Trizol, Sangon Biotech) according to the manufacturer's instructions. The quality of the total RNA samples was verified by the sequencing company. RNA was then enriched using magnetic beads containing oligo (dT). Fragmentation buffer was added to break the mRNA into

short fragments, which served as templates to synthesize single-stranded cDNA using random hexamers. Buffer, dNTPs, DNA polymerase I, and RNase H were then added for the synthesis of double-stranded cDNA, which was later purified using AMPure XP beads. The purified double-stranded cDNA was first subjected to end-repair, A-tailing, and adapter ligation, followed by size selection of the fragments using AMPure XP beads. Finally, polymerase chain reaction (PCR) amplification was performed; the products were purified using AMPure XP beads, and a final library was obtained.

Preliminary quantification of the constructed library was performed with Qubit2.0. The library was diluted to 1.5 ng/μl, and the insert size was determined using the Agilent 2100 system. After the expected insert size was obtained, accurate quantification of the effective concentration of the library was performed using quantitative PCR (valid effective concentration > 2 nM). This step was conducted to ensure the quality of the library, which was then subjected to sequencing.

2.3 Raw data processing and sequence stitching

Sequencing of the transcriptome library was performed using the Illumina HiSeq 4000 sequencing platform with the pair-end method (Novogene Bioinformatics Technology Co. Ltd., China). The raw data of the sequencing results (raw reads) with low-quality reads were removed. Low quality reads refer to those that has an adaptor N% > 0.1%, and those that consist more than 50% of the total reads which bases had a Qphred quality score ≤ 20. The clean reads were merged using Trinity software to obtain the unigenes of the sample (Grabherr *et al.*, 2011).

2.4 Functional annotation of *E. gracilis* unigenes

The resulting unigenes were annotated in seven large databases: Nr, Nt, Pfam (<http://pfam.sanger.ac.uk/>), KOG/COG (<http://www.ncbi.nlm.nih.gov/COG/>), Swiss-Prot (<http://www.ebi.ac.uk/uniprot/>), KEGG (<http://www.genome.jp/kegg/>), and GO (<http://www.geneontology.org/>).

3. Results

3.1 Illumina sequencing and de novo assembly

High-throughput sequencing data were used as the basis for transcriptome analyses of *E. gracilis*. The quality assessment of the sequencing data is shown in Table 1. A total of 622,740,910 raw reads were obtained, with a small proportion (3.77–4.70%) of reads containing adaptors, those of low quality, or those with undetermined sequences. After filtering the raw data, 595,483,638 clean reads were obtained. The total number of bases was 89.32 G, and the GC content was greater than 60% of the total number of bases. More than 96% and 91% of the reads showed a quality score of Q20 (base call accuracy of 99% or above) and Q30 (base call accuracy of 99.9% or above), respectively.

The clean reads merged using Trinity revealed a total of 297,648 transcripts and 120,086 unigenes. The lengths of the unigenes were ranged from 201 to 15,348 bp, with an average length of 636 bp and an N50 value of 1,059 bp (Table 2). The majority of the unigenes had a length of <301 bp,

Table 1. Summary of sequencing data for each sample

Sample	Raw reads number	Clean reads number	Clean bases	Q20 (%) ^a	Q30 (%) ^b	GC (%) ^c	Error (%) ^d
1	153,121,624	146227492	21.93 G	96.63	91.49	61.46	0.02
2	150,206,140	143,278,920	21.49 G	96.62	91.48	61.41	0.02
3	140,613,960	134,470,478	20.17 G	96.55	91.33	61.82	0.02
4	178,799,186	171,506,748	25.73 G	96.95	92.22	61.41	0.02
Total	622,740,910	595,483,638	89.32 G	-	-	-	-

Note: a Percentage of bases with Phred value N20 as a percentage of the total number of bases.

b Percentage of bases with Phred value N30 as a percentage of the total number of bases.

c Sum of G and C bases as a percentage of the total number of bases.

d Base error rate

Table 2. Length distribution of assembly

	Min length (bp)	Mean length (bp)	Median length (bp)	Max length (bp)	Total nucleotides	N50 ^a (bp)
Transcripts	201	726	446	15348	216,043,954	1117
Unigenes	201	636	341	15348	76,390,904	1059

Note: N50 Assembled transcripts were sorted from the largest to the smallest according to length; N50 is the transcript length when the length of the transcript is half of the total length.

accounting for 41.89% of all transcripts, followed by unigenes of 301–500 bp (24.56% of all transcripts). These data indicated that the sequencing results were highly accurate and could be used for further functional analyses.

3.2 Unigene function annotation and classification

3.2.1 Unigene sequence similarity

A total of 120,086 unigenes were assembled using Trinity software and were compared against seven large databases. A total of 48,031 unigenes were annotated, with data revealed 60% of the unigenes had no homologous sequences in the known databases. The results of the functional annotations of unigenes in the seven large databases are shown in Table 3. The largest number of transcripts were annotated in the GO database, and the least number of transcripts were annotated in the KEGG database and KOG databases.

The homology analysis of the unigenes among closely related species was based on the annotation results compared against the Nr database (Figure 1). These results showed that 31.85% of the *E. gracilis* unigenes had high sequence similarity to a gene in the Nr database from various species (Figure 1A), including *Oryza sativa japonica* (12.3%), *Guillardia theta* (6.3%), *Oryza sativa indica* (5.8%), *Emiliania huxleyi* (4.7%), and *Ectocarpus siliculosus* (2.7%). As shown in Figure 2B, 12.1% of the unigenes showed 95–100% homology, and 11.1% of the unigenes showed 80–95% homology with sequences in the database.

3.2.2 KOG annotation and classification

After KOG annotations, the successfully annotated unigenes were classified according to the three major KOG categories in the next level. As shown in Figure 2, the 11,436 unigenes annotated in the KOG database were divided into 26 major categories. The functional category with the largest

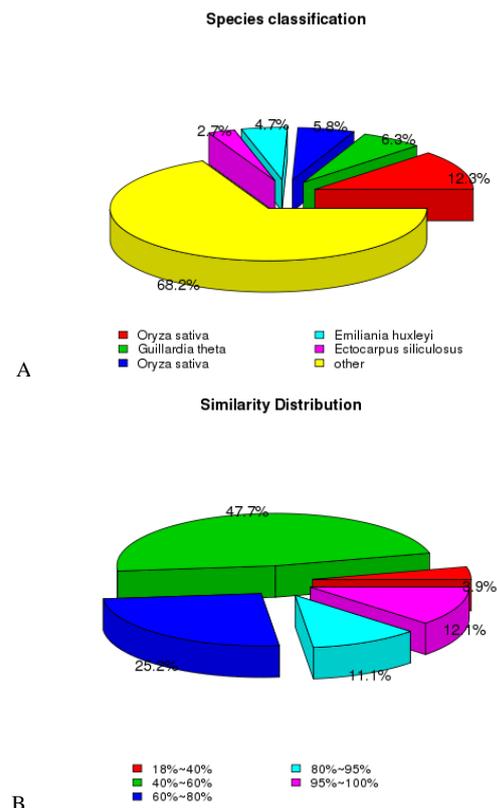


Figure 1. Statistics of NR annotation results, (A) Sequence of a species from the sample relative to near-source species (each sector represents a species). A larger fan-shaped area reflects a greater number of sequences aligned to the species. (B) Similarity refers to the ratio of similar amino acids to total amino acids (E-value < 1e-5). Each sector represents a similarity interval; a larger fan-shaped area reflects a greater number of genes with a similarity degree within this interval.

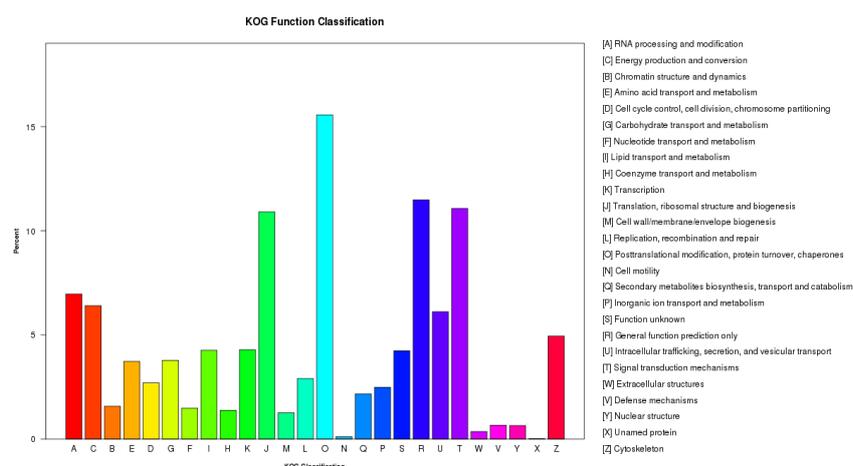


Figure 2. Statistics of KOG annotation results of unigenes, Axis-X represent the functional classification of a KOG (in uppercase letters A to Z; legend on the right referring for the specific meaning of each uppercase letters A to Z; Axis-Y indicates the number of unigenes with that function.

Table 3. Summary of the annotations of Unigenes

	Number of unigenes	Percentage (%)
Annotated in NR	24157	20.11
Annotated in NT	19367	16.12
Annotated in KEGG	9630	8.01
Annotated in SwissProt	23047	19.19
Annotated in PFAM	34131	28.42
Annotated in GO	34915	29.07
Annotated in KOG	11436	9.52
Annotated in all databases	3489	2.9
Annotated in at least one database	48031	39.99
Total unigenes	120086	100

number of unigenes was posttranslational modification, protein turnover, chaperones (category “O”), accounting for 13.96% of all KOG annotated unigenes, followed by general functional prediction (R, 10.30%); signal transduction mechanisms (T, 9.93%); and translation, ribosomal structure, and biogenesis (J, 9.78%).

3.2.3 GO annotation and classification

GO provides gene information and performs functional classification based on three major functions (molecular functions, biological process, cellular component). *E. gracilis* unigenes were annotated to GO terms using Blast 2.0 software. The 34,915 unigenes were classified into 56 functional categories (Figure 3), with the majority of genes (92,967) classified into biological process, followed by cellular component (54,713 genes) and molecular function (38,777 genes).

Within the biological process category, most transcripts were categorized to cellular process and metabolic process, accounting for 19,693 and 17,152 transcripts, respectively. Within the cellular component category, cell and cell part constituted the greatest number of transcripts. Within the molecular function category, the greatest number of transcripts were annotated to binding (17,865) and catalytic activity (14,379) functions.

3.2.4 KEGG pathway annotation

The genes with KEGG annotations were subjected to pathway analysis in the KEGG database to further identify the key metabolic pathways and signaling pathways involved, divided into five branches: cellular processes, environmental information processing, genetic information processing, metabolism, and organismal systems. A total of 9,630 unigenes were annotated into 128 KEGG pathways. Most of the annotated genes were included in metabolism (4,173), followed by genetic information processing (2,776). The majority of transcripts in the metabolism pathway were annotated to carbohydrate metabolism (812, 8.05%), and the majority of transcripts in the genetic information processing pathway were annotated to translation (1268, 12.56%) (Figure 4).

3.2.5 Unigenes related to starch and sucrose metabolism

A total of 114 unigenes were annotated to 29 enzymes involved in the starch and sucrose metabolism of *E. gracilis* (Table 4). Twelve unigenes were associated with glucose-6-phosphate isomerase (EC: 5.3.1.9), and seven genes were associated with UDP-glucuronate decarboxylase (EC: 4.1.1.35), phosphoglucomutase (EC: 5.4.2.2), and UDP-glucuronate 4-epimerase (EC: 5.1.3.6).

The pathway map of starch biosynthesis in *E. gracilis* was constructed based on the functional annotations of its transcriptome (Figure 5A). The major raw material for the synthesis of starch in *E. gracilis* was found to be adenosine diphosphate glucose (ADP-glucose), which is produced from glucose 1-phosphate (α -D-glucose-1P) under the catalytic action of glucose-1-phosphate adenylyl transferase (AGPase, EC: 2.7.7.27). ADP-glucose then produces amylose under the catalytic action of starch synthase (EC: 2.4.1.21), and the amylose is further synthesized into starch under the action of the 1,4- α -glucan branching enzyme (BE, EC: 2.4.1.18). Sucrose was involved in starch synthesis through three ways: Sucrose produces D-Fructose under

Table 4. Unigene related to starch and sucrose metabolism.

No.	KO name	Name	Unigene number	EC number
1	K01810	glucose-6-phosphate isomerase	12	5.3.1.9
2	K08678	UDP-glucuronate decarboxylase	7	4.1.1.35
3	K01835	phosphoglucomutase	7	5.4.2.2
4	K08679	UDP-glucuronate 4-epimerase	7	5.1.3.6
5	K01188	beta-glucosidase	6	3.2.1.21
6	K00688	starch phosphorylase	6	2.4.1.1
7	K00012	UDPglucose 6-dehydrogenase	6	1.1.1.22
8	K01193	beta-fructofuranosidase	6	3.2.1.26
9	K05350	beta-glucosidase	5	3.2.1.21
10	K01051	pectinesterase	5	3.1.1.11
11	K00975	glucose-1-phosphate adenylyltransferase	4	2.7.7.27
12	K16055	trehalose 6-phosphate synthase/phosphatase	4	2.4.1.15 3.1.3.12
13	K00700	1,4-alpha-glucan branching enzyme	4	2.4.1.18
14	K00847	fructokinase	4	2.7.1.4
15	K00963	UTP--glucose-1-phosphate uridylyltransferase	4	2.7.7.9
16	K01187	alpha-glucosidase	3	3.2.1.20
17	K13648	alpha-1,4-galacturonosyltransferase	3	2.4.1.43
18	K01179	endoglucanase	3	3.2.1.4
19	K00696	sucrose-phosphate synthase	3	2.4.1.14
20	K00703	starch synthase	2	2.4.1.21
21	K15920	beta-D-xylosidase 4	2	3.2.1.37
22	K05349	beta-glucosidase	2	3.2.1.21
23	K00844	hexokinase	2	2.7.1.1
24	K00695	sucrose synthase	2	2.4.1.13
25	K00845	glucokinase	1	2.7.1.2
26	K00705	4-alpha-glucanotransferase	1	2.4.1.25
27	K01177	beta-amylase	1	3.2.1.2
28	K01087	trehalose 6-phosphate phosphatase	1	3.1.3.12
29	K00770	1,4-beta-D-xylan synthase	1	2.4.2.24

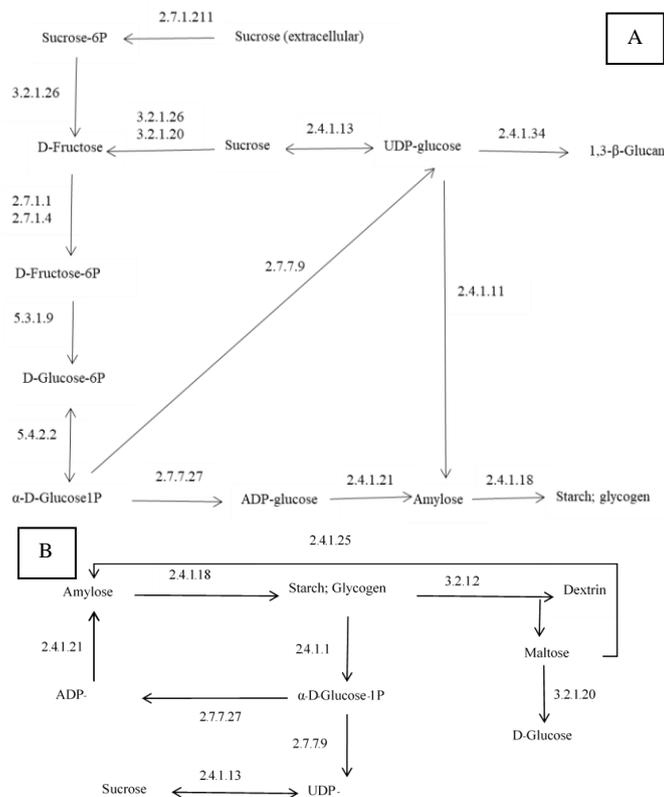


Figure 5. Construction of starch metabolic pathway in *E.gracilis* A, biosynthesis of the starch; B, catabolism of the starch

Table 5. Unigene for β -amy (EC:2.4.1.1) to match the gene in the gene bank database

Unigene	Length	Description	Genebank ID	Percent identity (%)	eValue
DN21867_c0_g1	874	Coccomyxa subellipsoidea C-169	545373473	54.08	1.20E-89
		Ostreococcus lucimarinus CCE9901	145352201	55.25	1.50E-87
		Auxenochlorella protothecoides	760449011	52.2	4.30E-87
		Ostreococcus tauri	308802832	53.74	9.50E-87
		Dunaliella salina	540198456	51.02	1.80E-85
		Micromonas pusilla CCMP1545	303282169	52.36	3.00E-85
		Micromonas sp. RCC299	255086713	53.04	2.00E-84
		Micromonas sp. RCC299	255085620	51.53	3.40E-84
		Galdieria sulphuraria	545703227	52.24	8.20E-10
		Galdieria sulphuraria	545703225	52.24	8.20E-10
DN52145_c0_g1	204	Galdieria sulphuraria	545701545	50	7.00E-09
		Thalassiosira oceanica	397566209	50.74	8.60E-93
		Thalassiosira pseudonana CCMP1335	224008156	48.08	5.60E-92
DN34998_c0_g1	2475	Phaeodactylum tricornutum CCAP 1055/1	219117381	49.26	4.40E-89
		Thalassiosira pseudonana CCMP1335	224007781	50.44	3.20E-87
		Phaeodactylum tricornutum CCAP 1055/1	219116725	49.7	7.10E-87
		Aureococcus anophagefferens	676389615	52.11	1.20E-86
		Aureococcus anophagefferens	676381895	51.76	2.70E-86
		Thalassiosira oceanica	397610212	51.59	1.30E-85
		Nannochloropsis gaditana CCMP526	553193611	49.25	1.70E-85
		Ectocarpus siliculosus	298715267	47.56	8.80E-70
		Aureococcus anophagefferens	676384469	47.42	1.20E-63
		Micromonas sp. RCC299	255071503	42.41	1.50E-56
DN29395_c0_g1	1237	Micromonas pusilla CCMP1545	303272831	41.18	2.60E-53
		Micromonas pusilla CCMP1545	303284159	44.76	2.10E-18
		Micromonas sp. RCC299	255086863	44.76	6.10E-18

4. Discussion

In the present study, a total of 48,031 unigenes were annotated in seven large databases, representing an annotation rate of 39.99% and 60.01% of the unigenes which do not have similar homologous sequences in the known databases. This may be due to the relatively deficient genetic sequence data for *Euglena* in public databases, or these unannotated unigenes may represent genes that are unique to *E. gracilis* (Chen, Li, Xiao, & Liu, 2015). The largest number of unigenes were annotated to the GO database, indicating predominant involvement in cellular process, metabolic process, cell, cell part, binding, and catalytic activity. There were 24,157 unigenes annotated in the Nr database, and a homology search revealed that *E. gracilis* was most closely related to *Oryza sativa* (Chen, Lam, & Chen, 1985). A large proportion of unigenes with KOG annotations were those involved in posttranslational modification, protein turnover, and chaperones; translation, ribosomal structure, and biogenesis; general functional prediction; and signal transduction mechanisms. This indicates the relatively high cellular activity of *E. gracilis*.

In the KEGG database, 114 unigenes were indicated to function in starch and sucrose metabolism, involving a total of 29 enzymatic genes. It was revealed that sucrose synthesis has priority for carbon partitioning, and starch synthesis is controlled by sucrose synthesis, but not vice versa (Stitt, M. *et al.*, 1987). Based on these findings, a pathway map of starch and sucrose biosynthesis in *E. gracilis* was constructed. Sucrose was indicated to be involved in the process of starch synthesis and decomposition, and the two are complementary. The enzymes involved in this pathway included phosphoglucomutase (EC: 5.4.2.2), which catalyzes the

conversion of glucose-6-phosphate to glucose-1-phosphate, and AGPase (EC: 2.7.7.27), which catalyzes the production of ADP-glucose and inorganic phosphate from glucose-1-phosphate and ATP. AGPase consists of two large subunits and two small subunits and can exist in two isoforms: cytosolic and chloroplast AGPase (Dawar *et al.*, 2013). AGPase is the major enzyme that regulates the key products of starch metabolism (Asenci³n Diez, Demonte, Guerrero, Ballicora, & Iglesias, 2013). Starch synthase (EC: 2.4.1.21) includes soluble starch synthase and granule-bound starch synthase which catalyzes the production of amylose from ADP-glucose, and upon activation by 1,4- α -glucan branching enzyme (EC: 2.4.1.18), forms the α -1,6 glycosidic bond in dextran, leading to starch production (Dian *et al.*, 2005; Leterrier *et al.*, 2008; Ma *et al.*, 2019). Thus, the present analysis showed that the starch synthesis rate of *E. gracilis* could be augmented by increasing the activity of AGPase (EC: 2.7.7.27) (Smith, 2010).

The results also showed that starch could be catabolized by both hydrolysis and phosphorylation, given that the transcripts of the enzymes involved in these two pathways were present in the transcriptome of *E. gracilis*. The key enzyme involved in the hydrolysis pathway was found to be β -amylase (β -amy, EC: 3.2.1.2), which catabolizes starch to form dextrin and maltose, and finally hydrolyzes maltose into d-glucose. Phosphorylase was involved in the phosphorylation pathway of starch production (Spase, EC: 2.4.1.1). This enzyme catalyzes the degradation of starch to α -d-glucose-1P. Two types of Spase enzymes were detected in *E. gracilis*: Spase1 and Spase2. Spase1 is located in the chloroplast, and they control starch synthesis by regulating the adjacent starch synthesis system (Liu *et al.*, 2014), whereas Spase2 located in the cytoplasm function in the catabolism of

starch (Irina *et al.*, 2014).

Given that β -amy (EC: 3.2.1.2) and Spase (EC: 2.4.1.1) are the key enzymes involved in the regulation of starch metabolism, knockout of the β -amy or Spase2 genes may be an effective strategy to reduce starch degradation, thereby improving the starch content in *E. gracilis*. Overall, these data served as a reference to further unveil the gene expression and regulatory mechanism of starch metabolism in *E. gracilis*. Outcomes generated from this study provided a scientific basis for the understanding of starch metabolic pathways in *E. gracilis* and they are potentially useful in the formulation of strategies to boost starch production for industrial application in the near future.

Transcriptome studies can infer the function of unknown genes based on the expression information of genes under specific conditions, and then reveal the mechanism of genes in the corresponding pathways. O'Neill *et al.*, (2015) studied the transcriptome of *E. gracilis* by Illumina sequencing technique, and obtained raw reads 383,414,636, clean bases 38.4 G. In the present study, the high-throughput transcriptional sequencing of samples from four experimental groups of *E. gracilis* was performed using Illumina platform, and raw reads 622,740,910 bp, clean bases 89.32 G were obtained. The number of reads obtained in this study is much higher than that obtained in O'Neill's study. In the study of O'Neill *et al.*, (2015), The paramylon is produced from UDP-glucose under the catalytic action of paramylon synthetase (EC:2.4.1.34), and then β -1,3-glucanase is produced under the catalytic action of Glucan-endo-1,3- β -glucosidase (EC:3.2.1.58). However, the final enzyme in paramylon synthesis (paramylon synthetase (EC 2.4.1.34)), was not found in any of the three transcriptomic libraries. We speculate that β -1,3-glucanase may be produced under the action of other enzymes, but this requires more evidence.

5. Conclusions

In this study, we used high-throughput sequencing technology to sequence the transcriptome of *E. gracilis* and obtained 297,648 transcripts and 120,086 unigenes. On the basis of functional annotation and cluster analysis of the transcriptome of *E. gracilis*, we preliminarily predicted the starch metabolism pathway and analyzed the key enzymes in this pathway. The starch metabolism pathway constructed by transcriptome sequencing is helpful for the genetic modification of *E. gracilis* and other microalgae to improve the synthesis and accumulation of microalgae polysaccharides.

Acknowledgements

This study was supported by the Tianjin Science and Technology Project (No. 18ZXYENC00070), and Tianjin Natural Science Foundation Project (No.18JCQNJC14800).

References

Afiukwa, C. A., & Ogbonna, J. C. (2017). Effects of mixed substrates on growth and vitamin production by *Euglena gracilis*. *African Journal of Biotechnology*, 26(22), 2612-2615. doi:10.5897/AJB2007.000-2417

- Ahmadinejad, N., Dagan, T., & Martin, W. (2007). Genome history in the symbiotic hybrid *Euglena gracilis*. *Gene*, 402(1), 35-39. doi:10.1016/j.gene.2007.07.023
- Armbrust, E. V., Berges, J. A., Bowler, C., Green, B. R., Martinez, D., Putnam, N. H., & Rokhsar, D. S. (2004). The genome of the diatom *Thalassiosira pseudonana*: Ecology, evolution, and metabolism. *Science*, 79-86. doi:10.1126/science.1101156
- Asenci n Diez, M. D., Demonte, A. M., Guerrero, S. A., Ballicora, M. A., & Iglesias, A. A. (2013). The ADP-glucose pyrophosphorylase from *Streptococcus mutans* provides evidence for the regulation of polysaccharide biosynthesis in Firmicutes. *Molecular Microbiology*, 90(5), 1011-1027. doi:10.1111/mmi.12413
- Chen, T. H., Lam, L., & Chen, S. C. (1985). Somatic embryogenesis and plant regeneration from cultured young inflorescences of *Oryza sativa* L. (rice). *Plant Cell Tissue and Organ Culture*, 4(1), 51-54. doi:10.1007/BF00041655
- Chen, X., Li, J., Xiao, S., & Liu, X. (2015). De novo assembly and characterization of foot transcriptome and microsatellite marker development for *Paspalum distachyon*. *Gene*, 576(1 Part 3), 537. doi:10.1016/j.gene.2015.11.001
- Cloonan, N., Forrest, A., Kolle, G., Ggardiner, B., Faulkner, G., Brown, M. K., . . . Bethel, G. (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nature Methods*, 5(7), 613-619. doi:10.1038/nmeth.1223
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., . . . Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), 644. doi:10.1038/nbt.1883
- Huang, Q. Y., Wang, Z. K., Hu, F., Zhang, C., & Lin, X. Z. (2018). Rna-seq based transcriptome analysis and the biosynthesis pathway of carotenoid in *Desmodesmus sp.* *Journal of Applied Oceanography*, 37(01), 68-76. doi:10.3969/J.ISSN.2095-4972.2018.01.008
- Irina, M., Sebastian, M., Saleh, A., Tom, O., Alisdair R. F., Otto, B., . . . Joerg, F. (2014). Double knockout mutants of *Arabidopsis* grown under normal conditions reveal that the plastidial phosphorylase isozyme participates in transitory starch metabolism. *Plant Physiology*, 164(2), 907-921. doi:10.1104/pp.113.227843
- Leterrier M., Holappa L. D., Broglie K. E., & Beckles K. E. (2008). Cloning, characterisation and comparative analysis of a starch synthase IV gene in wheat: Functional and evolutionary implications. *BMC Plant Biology*, 8(1), 98. doi:10.1186/1471-2229-8-98
- Li, H., Wang, W., Wang, Z., Lin, X., Fang, Z., & Yang, L. (2016). De novo transcriptome analysis of carotenoid and polyunsaturated fatty acid metabolism in *Rhodomonas sp.* *Journal of Applied Phycology*, 28(3), 1649-1656. doi:10.1007/s10811-015-0703-5

- Liu, F., Makhmoudova, A., Lee, E. A., Wait, R., Emes, M. J., & Tetlow, I. J. (2014). The amylose extender mutant of maize conditions novel protein-protein interactions between starch biosynthetic enzymes in amyloplasts. *Plant Physiology and Biochemistry*, 83(15), 4423-4440.
- Ma P., Yuan Y., Shen Q., Jiang Q., Hua X., Zhang Q., . . . Zhang J. (2019). Evolution and expression analysis of starch synthase gene families in *Saccharum spontaneum*. *Tropical Plant Biology*, 12, 158-173. doi:10.1007/s12042-019-09225-3
- Muriira, N. G., Xu, W., Muchugi, A., Xu, J. C., & Liu, A. Z. (2015). De novo sequencing and assembly analysis of transcriptome in the Sodom apple (*Calotropis gigantea*). *Bmc Genomics*, 16(1), 723. doi:10.1186/s12864-015-1908-3
- Nijkamp, J. F., Broek, M. V. D., Datema, E., Kok, S. D., Bosman, L., Luttkik, M. A., . . . Daran, J. M. (2012). De novo sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN.PK113-7D, a model for modern industrial biotechnology. *Microbial Cell Factories*, 11(1), 36-36. doi:10.1186/1475-2859-11-36
- Ogbonna, J. C. (2009). Microbiological production of tocopherols: Current state and prospects. *Applied Microbiology and Biotechnology*, 84(2), 217-225. doi:10.1007/s00253-009-2104-7
- O'Neill, E. C., Trick, M., Hill, L., Rejzek, M., Dusi, R., Hamilton, C. J., . . . Field, R. A. (2015). The transcriptome of *Euglena gracilis* reveals unexpected metabolic capabilities for carbohydrate and natural product biochemistry. *Molecular Biosystems*, 11(10), 2808-2820. doi:10.1039/c5mb00319a
- Pringsheim, E. G., & Hovasse, R. (2010). The loss of chromotophores in *euglena gracilis*. *New Phytologist*, 47(1), 52-87. doi:10.1111/j.1469-8137.1948.tb05092.x
- Pulz, O., & Gross, W. (2004). Valuable products from biotechnology of microalgae. *Applied Microbiology and Biotechnology*, 65(6), 635-648. doi:10.1007/s00253-004-1647-x
- Rismani-Yazdi, H., Haznedaroglu, B. Z., Bibby, K., & Peccia, J. (2011). Transcriptome sequencing and annotation of the microalgae *Dunaliella tertiolecta*: Pathway description and gene discovery for production of next-generation biofuels. *Bmc Genomics*, 12(1), 148. doi:10.1186/1471-2164-12-148
- Russo, R., Barsanti, L., Evangelista, V., Frassanito, A. M., Longo, V., Pucci, L., . . . Gualtieri, P. (2016). *Euglena gracilis* paramylon activates human lymphocytes by upregulating pro-inflammatory factors. *Food Science and Nutrition*, 5(2), 205. doi:10.1002/fsn3.383
- Sharkey, T. D., Stitt, M., Heineke, D., Gerhardt, R., Raschke, K., & Heldt, H. W. (1986). Limitation of Photosynthesis by Carbon Metabolism: II. O₂-Insensitive CO₂ uptake results from limitation of triose phosphate utilization. *Plant Physiology*, 81(4), 1123-1129. doi:10.1104/pp.81.4.1115
- Smith, A. M. (2010). Prospects for increasing starch and sucrose yields for bioethanol production. *Plant Journal*, 54(4), 546-558. doi:10.1111/j.1365313X.2008.03468.x
- Stitt, M., Gerhardt, R., Wilke, I., Heldt, H. W. (1987). The contribution of fructose-2,6-bisphosphate to the regulation of sucrose synthesis during photosynthesis. *Physiologia Plantarum*, 69, 377-386. doi:10.1111/j.1399-3054.1987.tb04304.x
- Takeyama, H., Kanamaru, A., Yoshino, Y., Kakuta, H., Kawamura, Y., & Matsunaga, T. (1997). Production of antioxidant vitamins, β-carotene, vitamin c, and vitamin e, by two-step culture of *Euglena gracilis* Z. *Biotechnology and Bioengineering*, 53(2), 185-190. doi:10.1002/(SICI)1097-0290(19970120)53:23.3.CO;2-6
- Ugrappa, Nagalakshmi, Zhong, Wang, Karl, Waern, . . . Snyder. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, 320(5881), 1344-1349. doi:10.1126/science.1158441
- Weimin D., Huawu J., & Ping W. (2005). Evolution and expression analysis of starch synthase III and IV in rice[J]. *Journal of Experimental Botany*, 2005, 56(412), 623-632. doi:10.1093/jxb/eri065
- Yan, K., Qin, W., Feng, R. Z., Zhou, W. H., & Fang, C. (2017). Transcriptome analysis of *Cinnamomum longepaniculatum* by high-throughput sequencing. *Electronic Journal of Biotechnology*, 28(C), S0717345817300283. doi:10.1016/j.ejbt.2017.05.006
- Yoshida, Y., Tomiyama, T., Maruta, T., Ishikawa, T., & Arakawa, K. (2016). De novo assembly and comparative transcriptome analysis of *Euglena gracilis* in response to anaerobic conditions. *Bmc Genomics*, 17(1), 1-10. doi:10.1186/s12864-016-2540-6
- Yu, M. J., Yang, S. J., & Lin, X. Z. (2016). De-novo assembly and characterization of *Chlorella minutissima* UTEX2341 transcriptome by paired-end sequencing and the identification of genes related to the biosynthesis of lipids for biodiesel. *Marine Genomics*, 25, 69-74. doi:10.1016/j.margen.2015.11.005
- Zhang, W., Gao, J., Zhou, W., & Yeong, Y. S. (2020). Transcriptome-based analysis of *euglena gracilis* lipid metabolic pathways under light stress. *Turkish Journal of Fisheries and Aquatic Sciences*, 20(6), 453-465. doi:10.4194/1303-2712-v20_6_04
- Zhu, S. Q., Gong, Y. F., Hang, Y. Q., Liu, H., & Wang, H. Y. (2015). Transcriptome analysis of *Dunaliella viridis*. *Hereditas*, 37(8), 828. doi:10.16288/j.yczs.15-032